International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

A Short Review on Various Aspects of Gene Sequencing and Metagenomics to Study Human Gut Microbiome

Anirban Goutam Mukherjee[1], Uddesh Wanjari[2]

[1]M.Sc Biotechnology (third semester), Vellore Institute of Technology, Vellore, Tamil Nadu, India, 632014
[2]M.Sc Biochemistry (third semester), Kamla Nehru Mahavidyalaya, Nagpur, Maharashtra, India, 440024

Abstract - *Metagenomics is a word which we commonly hear today. This word has become a part of our day to day lives specifically for a Microbiologist and a Biotechnologist and is not restricted to only these fields but also for various other disciplines. In simple words, metagenomics can be defined as the study of metagenome, actually this means the basic study of the genetic material which is obtained directly from the environmental samples such as soil, water, faeces. This process deals with the genomic analysis of microbial DNA directly from the communities present in the samples obtained directly from microbial communities.*

Gut microbiota commonly refers to the microorganisms that live in the digestive tract of human beings particularly in the intestine. In humans the gut microbiota comprises of the large number of bacterial species. This number is very large as compared to other body parts. There is often a mutual relationship between the human beings and the gut microflora .This can be established because some human gut organisms benefit the host by fermenting dietary fibers into short chain fatty acids such as acetic acid and butyric acid and these can be easily absorbed by the host. Intestinal bacteria also play an important role in metabolizing sterols, bile acids and xenobiotic compounds. They are also known for producing vitamin B and vitamin K. Therefore it becomes very important to understand the nature of gut microflora as they play a very vital role in a person's life.

Key Words: Metagenome, microbiota, xenobiotics, gut microflora, gene.

1. A COMPLETE REVIEW

The microorganisms present in the human gut are known to perform a wide variety of functions. One of the most common is the production of vitamins by the microorganisms [1]. Now these vitamins are absorbed by the host's large intestine [2,3].But it is very difficult to understand the complex relationship between the species and vitamin pathway [4,5] . Here the main focus is to understand the contribution of genes that are involved in transport mechanisms and vitamin biosynthetic pathways and how it is related to the corresponding species' abundance. The main methodology which can be applied is shotgun genome sequencing [1].This technique was performed with faecal metagenomes of individuals from four different countries. A study of eight B vitamin [6,7] and menaquinone are done .The biosynthetic and transporter genes in healthy subjects from different populations is compared. The whole genome metagenomic sequences are used and a continuous bioinformatic analysis is done. Metagenomic sequences are subjected to the Illumina paired end sequencing and these were obtained from European Nucleotide Archive at EMBL-EBI under the accession number PRJEB2054, PRJNA422434, PRJNA275349, and PRJNA389280 [8,9,10].

Results demonstrate high positive correlations between the species having the genes related to vitamin biosynthetic pathways and transporter mechanisms than that with each of them alone .This clearly indicates that these particular genes are widely distributed among the dominant phyla of the gut species. It is also observed that the total gene abundances remained constant across healthy populations at the global level. But critical analysis from the meta-transcriptomics data explains that the production and utilization potential of these enzymes is very complex and is not merely any direct linear relationship. Therefore it can be said that the species composition and their association with complex metabolic pathway related genes determine the functional genetic content of vitamin metabolism and it's abundance [11].

It is a proven idea that the gut microbiota is commonly associated with a large number of diseases [12]. The reason for these is widely contributed by many factors. Some of them can include climate, geography, host nutrition, lifestyle and medication [13,14,15]. Therefore it becomes very essential to gain knowledge about the various populations and their habits to get a better understanding and concept of their microbiome .A metagenomic study of intestinal microbiota from the Kazakh donors was conducted. This experiment involved 84 subjects, which includes both male and female, healthy subjects and also includes metabolic syndrome (MetS) patients aged 25–75 years, from the Kazakh administrative centre, Astana [21]. Characterization and description of these microbiomes is done in comparison with a global dataset(832 individuals from five countries on three continents).The correlations between microbiota and various parameters like nutritional data from food frequency questionnaires were explored [16,17,18]. This process involved the extraction of

total DNA from various faecal samples followed by the sequencing of the samples at the EMBL Gene Core facility using an illumine Hi Seq 2500.Around 2.7±1.1 Gbp of 100 base pairs (bp) paired-end shotgun sequencing reads was generated for each sample. This research provides a very interesting result [19,20,21].

Results indicate that the Kazakh microbiomes are very much different when compared both from European and East Asian counterparts. Interestingly they are found to be similar to other Cental asian microbiomes. The main difference is that the most samples falls within the Prevotella rich enterotype. Now this is a clear indication that this is a result of potential regional diet and lifestyle. Also the demonstration of the enterotype designation is observed which explains the stability within an individual over time in about 82% of cases. Various gut microbiome features are also observed that helps to distinguish MetS patients from controls [21].

Enterotype	Kazakh (%)	MetaHIT(%)	Kazakh(N)	MetaHIT(N)
Bacteroidetes-rich	6	27	5	74
Prevotella-rich	71	26	60	71
Firmicutes-rich	23	48	19	133

(The table shows enterotype assignment counts for the novel Kazakh samples as well as Danish Meta HIT samples for comparisons, shown as sample counts (N) and as percentages(%).[21])

Exo-polysaccharides are commonly known polymers .These high molecular weight molecules are structurally and functionally valuable polymers secreted by the microorganisms under stress conditions. Such molecules are widely produced by bacterial species and are commonly found in faeces [22] .These inhibit DNA restriction and often cause problems in metagenomic analysis [23,24,25]. Therefore an approach was taken to determine the effects of different DNA extraction methods on the composition of gut microbiota using Illumina Mi Seq deep sequencing technology. This process included the extraction of DNA from the stool of an obese female using 10 different methods and the choice of the method that was employed for DNA extraction affected the abundance at the phylum level, diversity and species richness. Additional DNA was obtained from the stools of 83 different individuals. This involved the fastest extraction assay that degraded exo polysaccharides. This study shows the 16S rRNA [26-30] gene sequencing of the faecal microbial community of an obese subject using 10 different DNA extraction protocols. A total of 1,443,537 high quality sequences of around 411 bp in length were analyzed. It is found that the operational taxonomic unit (OTU) [31] for all the sample reads was found to be between167 to 2,127 for each extraction method. The OTU s found is demonstrated by the rarefaction curves. This provides a relation between sequencing effort and the number of putative species in the samples. The highest number of reads is found to be 465,988 and the lowest number to be 28,277 respectively [32]. The analysis was halted for every extraction method whenever the number of OTU s reached a plateau. The analysis produced a highest number of OTUs per number of reads with extraction method 5 whereas in contrast a high number of OTUs with extraction methods 9 and 10 was also obtained but this analyzed a comparatively large number of samples [33]. Additionally lower number of OTUs were found with various other extraction methods particularly with the method 8 resulting in the lowest species diversity [33].

In modern day the term colorectal cancer is gaining importance day by day and the number of cases of colorectal cancer is also increasing [34].Colorectal cancer is a cancer that initiates from the colon or the rectum. The cancer begins as a growth on the inner lining of the colon or rectum called polyps [35]. It is often taken as a case of molecular profiling. Now the association and contribution of bacterial species to colorectal cancer (CRC) is studied and they are related to consensus molecular subtypes (CMS) [36].This provides a better understanding of the relationship between the bacterial species and the various molecular mechanisms associated with CRC subtypes. In this process ,a classification of 34 tumours were done into CRC subtypes.RNA sequencing derived gene expression and determined relative abundances of bacterial taxonomic groups using 16S rRNA [38] amplicon meta barcoding. Sample preparation, library creation and ribosomal RNA depletion was performed using Illumina TruSeq V2 reagents and the proper RNA sequencing was carried out with Illumina HiSeq. 2500 V4 platform. This was done to produce 125 bp long paired end reads. Also a new approach was demonstrated .This showed libraries were sequenced using two lanes in which each sample library was split equally to the two lanes. Also the two lanes were merged. The libraries of 16rRNA were prepared. This is done using 20ng of DNA for each sample and by the utilization of primer pairs flanking the V3 and V4 regions of the 16S rRNA gene (16SFV35′-TATGGTAATTGGCCTACGGGAGGCAGCAG3′, 16SR_V4:(5′AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3′) and Illumina sequencing [38,39] adaptors and barcodes added using limited cycle PCR (40 cycles).Another important procedure called Amplicon sequencing is carried out by using Illumina MiSeq platform, and paired end reads of length 250 bp were generated [40]. This research demonstrates the distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Classification of 34 tumours into colorectal cancer (CRC) was done [38] .The process was done by 16S rRNA amplicon metabarcoding.16S rRNA

analysis demonstrated the decreased or very reduced levels of Firmicutes and Proteobacteria in CMS1(tumour consensus molecular subtypes) and enrichment or very high levels of Fusobacteria and Bacteroidetes was observed. Additionally an analysis of bacterial taxa using non human RNA sequencing reads was also done and this exhibited distinct bacterial communities associated with each molecular subtype. The most highly enriched species which were found to be associated with CMS1 included mainly *Fusobacterium hwasookii* and Porphyromonas gingivalis [41]. On the other hand CMS 2 was found to be enriched with *Selenomas* and *Prevotella* species, while CMS3 had few significant associations. These findings were validated by PCR and also proved the high enrichment of *Fusobacterium nucleatum*, *Parvimonas micra* and *Peptostreptococcus stomatis* in CMS1 [38].

The gut microbiome plays a crucial role within the various gastrointestinal diseases. Irritable bowel syndrome (IBS) and Inflammatory bowel disease (IBD) are most ordinarily known examples. Therefore the gut microbiome can exhibit a dual role both as a diagnostic tool and as a target for treatment. This study bridges the gap between functional studies and 16S.This is done by identifying complete species and functional gut microbiome profiles and therefore the purpose is served by high resolution shotgun metagenomic sequencing. An outsized case control study was performed comprising stool samples of 1792 individuals :355 IBD [42] patients (comprising patients with Crohn's disease (CD) and colitis (UC)), 421 IBS patients and 1025 population controls. This is often followed by isolation and sequencing of faecal DNA samples and this generated about 3.0Gb of knowledge per sample. Also an outsized number of phenotypes were collected for all participants. This included lifestyle information and disease characteristics. The most aim of the study indicates the inference of taxonomy, bacterial strain diversity, growth rates, gene families, virulence factors and antibiotic resistance protein abundances from the sequenced data. This research is completed to analyse gut metagenomes which reveals microbial treatment targets for inflammatory bowel disease4 and irritable bowel syndrome. An in depth multi-layer gut microbiome profiles were uncovered for both IBD43 and IBS.A substantial overlap in microbial species in patients with IBD and IBS were compared with controls. The overlap exhibited a decreased abundance in *Faecalibacterium prauznitzii* abundance in both IBD and IBS [43]. But a completely unexpected result's observed just in case of Crohn's disease which showed not only the abundance but also the strain diversity and also the amount of Escherichia coli number was increased [42]. These gut microbial properties successfully distinguished IBD and IBS. It's observed that the varied bacterial virulence factors involved in mucosal damage and iron uptake were found to be higher in patients with IBD [44]. Not only this but also the antibiotic resistances proteins, especially efflux pumps, were increased in both diseases .1-arginine deficiency are often explained by large alterations in microbial metabolic functions in CD disease. This will be seen in wound healing and vitamin B2 depletion [42]. This study basically explains the microbial diseases which will be distinguished between these gastrointestinal diseases.

2. GAPS AND ALTERNATIVE MEASURES

Often a variety of problems are encountered with such metagenomic studies. Generic annotation is often incomplete and sometimes misunderstood. One major problem is that despite the event of the many procedures, indicators and genetic tools, still lack effective screening methods for several activities. Additionally the inefficient expression of some metagenomic[20] genes within the host bacteria used for screening. Many metagenomic genes [19] are often derived from bacteria that have highly divergent physiologies and organic phenomenon machineries. These are generally absent from the surrogate host. Experimental design and procedures can directly affect the detected composition of the faecal bacterial community, including sampling and storage protocol also as DNA extraction method. Faeces represent one among the foremost complex biological materials for bacterial DNA isolation. This is often because it contains remnants of human DNA, food DNA and also many inhibitors .This may hamper and make problems in subsequent PCR amplification and NGS procedures. But these gaps are currently being taken into consideration and continuous research is being undertaken to fill these gaps. Now the varied newest generation of sequences like PacBio RS, the Ion Torrent Proton or the ONT GRIDIon/MINIon, will still propel the sector of metagenomics. Various innovative methods just like the affordable bench-top devices (454 Junior, Illumina MiSeq[29], Ion Torrent PGM and Proton) has totally changed the concept of sequencing, and future devices like the ONT MINIon also can be other alternatives. Illumina and Thermo platforms with their MiSeq and Ion Torrent (IT) Personal Genome Machine (PGM) benchtop sequencers have proved to be highly efficient for 16S rRNA-based[30] analyses of diverse bacterial populations.

3. CONCLUSIONS

The gut microbiota plays a good sort of roles. Therefore it becomes very essential to know to varied mechanisms of their functioning .These research studies demonstrates the interrelationship between B-Vitamin dosage on gut microbe and therefore the host and also the effect on the microbiota with under or over nutrition. This review also describes gut microbiome data from an understudied population. This will guide us for further comparative work on biogeography and research on widespread diseases. Additionally this study provides a stimulating incontrovertible fact that DNA are often obtained from exo polysaccharides commonly present in stool samples which may be a superb source for the metagenomic studies. Not

only this, the study also demonstrates the connection between the human gut microbiota with colorectal cancer, Irritable Bowel Syndrome and Inflammatory Bowel Disease. This review aims to provide us with some raw idea about the future scientific methods and techniques which would be highly efficient in the study of metagenome. Also this would further guide us to understand the complex relationship between the human beings and the gut microbiota.

REFERENCES

[1] Said HM, Kumar C. Intestinal absorption of vitamins. Curr Opin Gastroenterol. 1999;15:172–6.

[2] Said HM, Mohammed ZM. Intestinal absorption of water-soluble vitamins:an update. Curr Opin Gastroenterol. 2006;22:140–6.

[3] Degnan PH, Taga ME, Goodman AL. Vitamin B12 as a modulator of gut microbial ecology. Cell Metab. 2014;20:769–78.

[4] Said HM, Mohammed ZM. Intestinal absorption of water-soluble vitamins:an update. Curr Opin Gastroenterol. 2006;22:140–6.

[5] Said HM, Nexo E. Gastrointestinal Handling of Water-Soluble Vitamins.Compr Physiol. 2018;8:1291–311

[6] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55–60.

[7] Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et.al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol. 2018;3:337–46.

[8] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing.Nature. 2010;464:59–65.

[9] Arumugam M, et al. Enterotypes of the human gut microbiome. Nature.2011;473:174–80.

[10] Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al.Richness of human gut microbiome correlates with metabolic markers.Karine Clément. 2013;13:28.

[11] Das, P., Babaei, P., & Nielsen, J. (2019). Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome. BMC Genomics, 20(1).

[12] Bakker GJ, Nieuwdorp M. Chapter 29 - Relationship Between Gut Microbiota, Energy Metabolism, and Obesity A2 - Floch, Martin H. In: Ringel Y, Walker WA, eds. The microbiota in gastrointestinal pathophysiology. Boston: Academic Press, 2017:255–8.

[13] Eren AM, Sogin ML, Morrison HG, et al. A single genus in the gut microbiome reflects host preference and specificity. Isme J 2015;9:90–100.

[14] Shanahan F, van Sinderen D, O'Toole PW, et al. Feeding the microbiota: transducer of nutrient signals for the host. Gut 2017;66:1709–17.

[15] Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. Biol Lett 2014;10:20131037.

[16] Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for earlystage detection of colorectal cancer. Mol Syst Biol 2014;10:766.

[17] Kultima JR, Sunagawa S, Li J, et al. MOCAT: a metagenomics assembly and gene prediction toolkit. PLoS One 2012;7:e47656.

[18] Sunagawa S, Mende DR, Zeller G, et al. Metagenomic species profiling using universal phylogenetic marker genes. Nat Methods 2013;10:1196–9.

[19] Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 2014;32:834–41.

[20] Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 2012;490:55–60.

[21] Kushugulova, A., Forslund, S. K., Costea, P. I., Kozhakhmetov, S., Khassenbekova, Z., Urazova, M., Bork, P. (2018). Metagenomic analysis of gut microbial communities from a Central Asian population. BMJ Open, 8(7), e021682.

[22] Angelakis, E. et al. A Metagenomic Investigation of the Duodenal Microbiota Reveals Links with Obesity. PloS One. 10, e0137784(2015).

[23] Angelakis, E., Armougom, F., Million, M. & Raoult, D. The relationship between gut microbiota and weight gain in humans. Future.Microbiol 7, 91–109 (2012).

[24] Angelakis, E., Merhej, V. & Raoult, D. Related actions of probiotics and antibiotics on gut microbiota and weight modification.Lancet Infect Dis 13, 889–899 (2013).

[25] Mirsepasi, H. et al. Microbial diversity in fecal samples depends on DNA extraction method: easyMag DNA extraction compared to QIAamp DNA stool mini kit extraction. BMC. Res. Notes 7, 50 (2014).

[26] Henderson, G. et al. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PloS One. 8, e74787 (2013).

[27] McOrist, A. L., Jackson, M. & Bird, A. R. A comparison of five methods for extraction of bacterial DNA from human faecal samples.J. Microbiol Methods 50, 131–139 (2002).

[28] Smith, B., Li, N., Andersen, A. S., Slotved, H. C. & Krogfelt, K. A. Optimising bacterial DNA extraction from faecal samples:comparison of three methods. Open. Microbiol J. 5, 14–17 (2011).

[29] Vo, A. T. & Jedlicka, J. A. Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. Mol. Ecol. Resour. 14, 1183–1197 (2014).

[30] Milani, C. et al. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PloS One. 8, e68739(2013).

[31] Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol 75, 7537–7541 (2009).

[32] Zoetendal, E. G. et al. Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nat. Protoc. 1, 954–959 (2006).

[33] Angelakis, E., Bachar, D., Henrissat, B., Armougom, F., Audoly, G., Lagier, J.-C., Raoult, D. (2016). Glycans affect DNA extraction and induce substantial differences in gut metagenomic studies. Scientific Reports, 6(1).

[34] Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136, E359–386.

[35] Jass, J. R. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology50, 113–130.

[36] Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. Nat Med 21, 1350–1356.

[37] Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27,2957–2963.

[38] Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335–336.

[39] Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S., & Frizelle, F. A. (2017). Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Scientific Reports, 7(1).

[40] Vogtmann, E. et al. Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing.PLoS One 11, e0155362.

[41] Binder Gallimidi, A. et al. Periodontal pathogens Porphyromonas gingivalis and Fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model. Oncotarget 6, 22613–22623.

[42] Vich Vila, A., Imhann, F., Collij, V., Jankipersadsing, S. A., Gurry, T., Mujagic, Z., Weersma, R. K. (2018). OP014 Analysis of 1792 gut metagenomes reveals microbial treatment targets for inflammatory bowel disease and irritable bowel syndrome. Journal of Crohn's and Colitis, 12(supplement_1), S010–S010.

[43] Peterson, D. A., Frank, D. N., Pace, N. R., & Gordon, J. I. (2008). Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. Cell Host & Microbe, 3(6), 417–427.

[44] Sartor, R.B. (2008). Microbial influences in inflammatory bowel diseases. Gastroenterology 134, 577–594